


Fold-change threshold screening: a robust algorithm to unmask hidden gene expression patterns in noisy aggregated transcriptome data

Jonas Hausen¹  · Jens C. Otte² · Uwe Strähle² · Monika Hammers-Wirtz³ · Henner Hollert¹ · Steffen H. Keiter^{1,4} · Richard Ottermanns¹

Received: 31 July 2014 / Accepted: 2 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Transcriptomics is often used to investigate changes in an organism's genetic response to environmental contamination. Data noise can mask the effects of contaminants making it difficult to detect responding genes. Because the number of genes which are found differentially expressed in transcriptome data is often very large, algorithms are needed to reduce the number down to a few robust discriminative genes. We present an algorithm for aggregated analysis of transcriptome data which uses multiple fold-change thresholds (threshold screening) and p values from Bayesian generalized linear model in order to assess the robustness of a gene as a potential indicator for the treatments tested. The algorithm provides a robustness indicator (ROBI) as well as a significance profile, which can be used to assess the statistical significance of a given gene for different fold-change thresholds. Using ROBI, eight discriminative genes were identified from an exemplary dataset (*Danio rerio* FET treated with chlorpyrifos, methylmercury, and PCB) which could be potential indicators for a given substance. Significance profiles uncovered genetic effects and revealed appropriate fold-change

thresholds for single genes or gene clusters. Fold-change threshold screening is a powerful tool for dimensionality reduction and feature selection in transcriptome data, as it effectively reduces the number of detected genes suitable for environmental monitoring. In addition, it is able to unmask patterns in altered genetic expression hidden by data noise and reduces the chance of type II errors, e.g., in environmental screening.

Keywords Bioinformatics · Masked effects · *Danio rerio* · Aggregated analysis · Ecotoxicogenomics · Robustness indicator (ROBI) · Bayesian generalized linear model

Introduction

Transcriptomics can be used in environmental screening for detecting an impact of substances on individuals and populations from altered gene expression. However, to enable the full potential of this tool, it is necessary to develop statistical methods for the detection of indicating patterns and discriminative genes for pollutants from gene expression data. Recently, very specific assessment methods for aquatic organisms have been developed in order to screen for pollutants and characterize their effects, esp. on zebrafish (*Danio rerio*) (Hollert et al. 2003; Höss et al. 2010; Feiler et al. 2013). Zebrafish embryos have become a useful model organism in ecotoxicology and ecotoxicogenomics (Yang et al. 2007; Strähle et al. 2012; Busquet et al. 2014; Schiwy et al. 2014). Recent research focused on linking gene expression analysis to current assessment methods for zebrafish embryos making it possible to investigate mechanism-specific toxicity at molecular level (Keiter et al. 2010).

Responsible editor: Philippe Garrigues

✉ Jonas Hausen
jhausen@bio5.rwth-aachen.de

- ¹ Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany
- ² Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
- ³ Research Institute for Ecosystem Analysis and Assessment - gaiaac, Kackertstraße 10, 52072 Aachen, Germany
- ⁴ Man-Technology-Environment Research Centre, Örebro University, 701 82 Örebro, Sweden

Detecting genes that show a different expression in comparison to a control sample might be the most basic yet most important step in this task as it allows identifying genes that respond to a special contamination (Kosmehl et al. 2012). In ecotoxicology, the challenge of taking the sources of variation in gene expression into account is an important task, e.g., due to genetic polymorphism in populations (Denslow et al. 2007). Oleksiak et al. (2002) analyzed genetic variation among natural populations in several studies, showing that up to 20 % of the tested genes showed statistically significant differences in expression at population level in similar environment.

When handling gene expression data in ecotoxicology, these sources of variability have to be taken into account as they can mask effects of the tested contaminants (Kerr et al. 2000; Hallare et al. 2011). Despite the large number of different technologies for gene expression analysis, data analysis and the consequent extraction of biological knowledge remain a major issue of transcriptomics as there is still no gold standard for transcriptome data analysis up to date (Cordero et al. 2007; Reboiro-Jato et al. 2014). This often makes the interpretation of gene expression data subjective, and there is need for tools to come to an objective conclusion (Yang et al. 2002). Algorithms with enough statistical power, the ability to unmask potential indicator genes and to unhide gene expression patterns, are needed to obtain biologically relevant results and reduce the chance for false-negative conclusions (type II error). In addition, indicator genes must be robust discriminators to show reproducible expression in multiple experiments (McCarthy and Smyth 2009).

The large number of tested genes in transcriptome data in comparison to the usually small sample size makes multiple testing one of the most dangerous pitfalls in the interpretation of statistical results (Breitling et al. 2004). Selecting discriminative genes is therefore a critical step for obtaining accurate and predictive information from gene expression data (Peng et al. 2014). An advantage of using statistical tests over fold-change criteria is that one gets a p value as a starting value for an assessment of the suitability of a gene. However, due to the large number of genes involved in transcriptome analysis applying p values is not always straightforward. On a microarray with 10,000 genes, a significance level of 0.05 would lead to 500 false-positive genes, mistakenly identified as significant (type I error). One obvious solution would be to consider a more stringent p value, e.g., by using a false discovery calculation (FDR; Benjamini et al. 2001). Unfortunately, this may result in excluding important genes, thereby increasing type II error. Furthermore, even with a more stringent p value, the number of discriminative genes provided by the statistical test of choice has to be further reduced to meet the requirements of practicability for an environmental screening approach. In conclusion, to assess the stability of a discriminative gene, statistical methods are required to reduce the chance of false-

positive detection even when sample size is small and the overall variation in the data is high due to data noise.

Possibilities to handle these problems are to include only genes with a fold-change above a certain threshold (double filtering; Zhang and Cao 2009) or to transform data into binary or ranked variables (Ding and Hanchuan 2005). However, this requires the selection of a cut-off value which is arbitrary without any background information. Similar issues arise for certain related gene expression analysis like double filtering, the usage of volcano plots, or functional analyses which depend on fold-change cut-off values.

The aim of this study was to develop an algorithm for feature selection and statistical pattern recognition in transcriptomic data. At its core, it is a multi-threshold approach based on binary variables to handle data noise and high variability. This algorithm is used to (1) select discriminative genes from multi-pollutant aggregated microarray data (thereby reducing the data dimensionality to a manageable set of variables), (2) assess the statistical stability of selected discriminative genes (giving an idea about uncertainty and indicative power), (3) detect hidden patterns in gene expression for target contaminants (unmasking effects superposed by data noise), and (4) reveal appropriate fold-change thresholds for each gene (enabling use of genes or gene patterns for prediction or classification).

Material and methods

Microarray data and analysis

Fertilized wild-type zebrafish eggs were exposed to three different treatments: methylmercury (mehg), chlorpyrifos (chlor), and Aroclor 1254 (PCB). As sediments are an important sink for a multitude of pollutants, the three substances were chosen as sediment pollutants often found in combination in environmental samples (Chapman 2000; Chapman and Anderson 2005). We used an aggregated approach of analysis to find discriminating genes in this multi-pollutant context. The aggregated approach makes it possible to select robust genes, which have a high discriminative power in the attempt to distinguish certain contaminants as distinguishing between specific responses and universal detoxification is often a problem (Snell et al. 2003; Fedorenkova and Vonk 2010; Kosmehl et al. 2012). The exposure to the chemicals was performed from 24 to 48 h post fertilization (hpf). To avoid gene-specific dye effects, microarrays were replicated with reverse-labeling to balance green and red dyes. The following processing of raw data, quality control, and analysis of variance was done according to the method given in Legradi (2011). Microarray preparation and analysis were part of the DanTox project (Keiter et al. 2010). For further details, see Hausen et al. (2015) and Keiter et al. (2013).

Fold-change threshold screening: the structure of the algorithm

The algorithm (see Fig. 1) is performed on microarray data of each gene in the dataset and was written in R (R Core Team 2014, Version 3.0.2).

Log₂ fold-changes are calculated as the logarithmic ratio of treatment and control spot intensity. To make sure that there are enough replicates for the following analysis, a missing value imputation of the fold-change values is applied. Only genes with at least two replicates for each treatment are retained. If a single replicate is missing for a treatment, the median is imputed. This way all genes that passed pre-selection have the same number of replicates.

The algorithm compares every fold-change with a series of predefined cut-off thresholds and creates a new binary 0/1 variable for each threshold. Fold-changes lower than the threshold (or higher in case of a negative threshold) are

transformed to state 0. Basically, the threshold can be seen as the cut-off beyond which each fold-change is considered significant (over- or under-expression). Instead of arbitrarily choosing a cut-off for all genes, fold-change threshold screening is designed as a step-down algorithm applying multiple cut-offs. The result of these steps is one binary variable for each threshold which will be used in further analysis.

Next, a Bayesian generalized linear model (function bayesglm from the arm package; Gelman et al. 2008) is performed on the binary data with treatment as predictor variable. In contrast to the default generalized linear model (glm), Bayesian generalized linear model uses a weakly informative prior distribution to handle small sample size as well as the perfect segregation problem which naturally occurs in this kind of data (Lesaffre and Albert 1989; Zorn 2005). To determine whether including the treatment as predictor has a significant effect on the model, models with and without predictor (null models) are compared using analysis of deviance tables (function anova.glm from the stats package). Subsequently, the next threshold is chosen as a cut-off value and the fold-changes are compared to it. Repeating this process for a number of thresholds, one gains a *p* value (from the analysis of deviance chi-squared test) for each of these thresholds.

In the next step, the number of thresholds which lead to a significant *p* value is counted. So, for each treatment, one then gets a number between 0 and the number of applied cut-off thresholds. This value is called the robustness indicator (ROBI). To be able to compare ROBI in different experiments, it can be normalized by dividing it by the number of applied cut-off thresholds so it ranges between 0 and 1. This value can be used to assess the robustness of the gene as a discriminator for one of the tested treatments. Robust genes indicate the same treatment despite changing thresholds. Analysis is also done with switched signs of all thresholds to test for significant downregulation of the gene.

To determine which treatment is causing a significant up-regulation or downregulation, a Tukey post-hoc test is performed on the Bayesian glm results. However, because of the small sample size, Tukey post-hoc test fails to find significant differences between the treatments even when the data are perfectly segregated between the treatment classes. Therefore, instead of performing post-hoc test on the result of each threshold separately, post-hoc test is calculated using the results of all thresholds. For each fold-change, the number of times the threshold is lower than the fold-change was counted resulting in a binomially distributed variable. With this variable, it is possible to perform a post-hoc test on the results of a Bayesian generalized linear model which uses the information from all thresholds.

The *p* values from anova.glm are plotted against the different thresholds used. The result is called a significance profile. Using the significance profile, it is possible to gain an

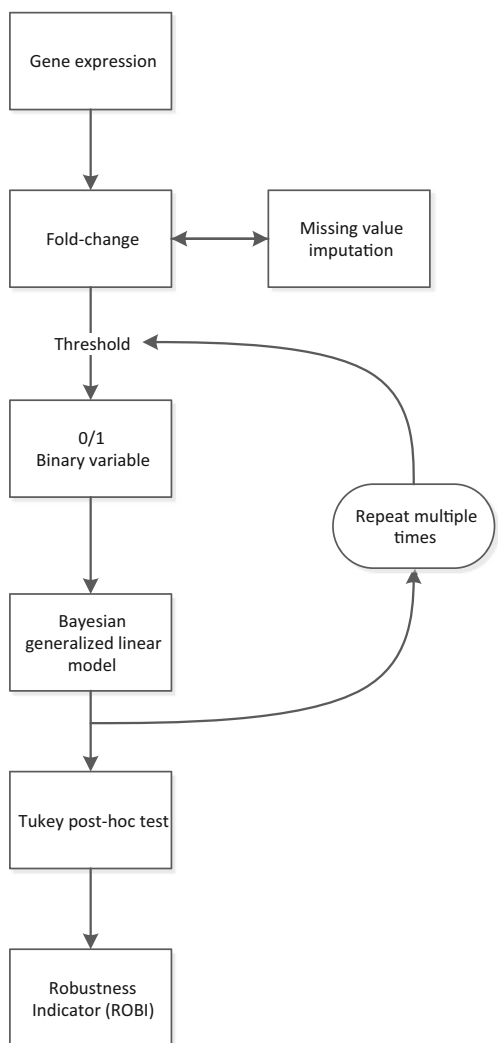


Fig. 1 Flowchart of fold-change threshold screening algorithm showing all steps performed

overview of the performance of the gene when different thresholds are applied, information which ROBI alone cannot provide.

To evaluate the results from discriminatory statistics and to verify the membership of extracted discriminative genes for gene regulation clusters, genes are clustered hierarchically based on Euclidean distance using Wards clustering method. Genes and treatments are displayed with distance relative to their relationship and differential expression illustrated by a color gradient heat map (Eisen et al. 1998) (heatmap.2 function from the R-package gplots).

Settings in the present study

In our study, we used the thresholds 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5 (and accordingly -0.5, -1, -1.5, -2, -2.5, -3, -3.5, -4, -4.5, and -5) as cut-off values. For the ROBI, a *p* value below 0.1 was counted once, a *p* value below 0.05 twice. This allowed the ROBI to range between 0 and 20. ROBI normalization, as described above, was performed additionally.

Results

From 4298, eight genes were identified as robust while highly significant discriminative for the treatments applied. Transformation according to the first threshold (0.5 or -0.5) done by the algorithm is exemplarily shown for these genes (Table 1).

For each of the three different chemicals, at least one gene was chosen as an example of a potential indicator gene. *p*-values from the anova.glm performed on the binary values are given in Table 2.

The robustness indicator (ROBI) was calculated by counting the number of significant thresholds once (*p*<0.1) and highly significant thresholds twice (*p*<0.05). As in this study, thresholds from 0.5 to 5 (and -0.5 to -5)

were used; the ROBI could range from 0 to 20. For example, for *mmp9*, the analysis provided six thresholds with highly significant *p* values ($\alpha < 0.05$) and one threshold with significant *p* values ($\alpha < 0.1$) so the ROBI was 15; the normalized ROBI would be 0.75.

For *mmp13a*, every applied threshold provided a highly significant *p* value (Fig. 2, left); whereas, *krt18* revealed highly significant effects for four thresholds (1.5 to 3) and insignificant results when another threshold was applied (Fig. 2, right).

Nine hundred-thirty genes with a mean fold-change >2 for at least one treatment were selected to calculate the heatmap (Fig. 3). The locations of the detected methylmercury indicators from Table 2 (*fn1b*, *socs3a*, *mmp13a*, *mmp9*, *sepw1*, *zgc:103438*) are marked on the right side. On the left side, corresponding methylmercury regulation clusters are marked. This allows detection of gene patterns that were significantly differently regulated in the presence of at least one of the different treatments (in this case methylmercury). The chosen genes that were upregulated form a tight cluster (“mehg +”) which corresponds well to the results of our algorithm.

Discussion

Often, analysis of gene expression data starts by creating and comparing fold-changes for each gene. However, because the use of fold-changes is no statistical test, there is no value which can indicate any level of confidence for the genes being differentially expressed or not (Pavlidis 2003; Breitling et al. 2004). Therefore, it is necessary to use a statistical test that is able to detect genes based not only on the fold-change but also on the reliability of reaching stable results without the use of a subjectively chosen cut-off threshold (McCarthy and Smyth 2009).

By using a glm-based approach, it is possible to reduce some of the drawbacks of such a fold-change-based approach

Table 1 Fold-changes of eight genes selected by fold-change threshold screening as well as results of the exemplary transformation with a threshold of 0.5

	cish	fn1b	socs3a	sepw1	mmp13a	mmp9	zgc:103438	itgb1b.2
chlor1	3.24/1	0.55/1	0.94/1	1.22/1	0.29/0	1.65/1	0.24/0	-0.83/1
chlor2	2.14/1	-1.22/0	-0.72/0	0.14/0	0.32/0	0.53/1	-0.51/0	-0.25/0
chlor3	2.69/1	-0.42/0	1.05/1	0.45/0	0.39/0	0.80/1	0.83/1	-0.43/0
mehg1	-0.59/0	6.48/1	10.93/1	-6.57/1	10.91/1	15.30/1	8.51/1	-0.91/1
mehg2	-0.13/0	5.54/1	5.79/1	-4.03/1	5.07/1	8.53/1	7.2/1	-0.71/1
mehg3	-0.86/0	6.12/1	10.51/1	-6.73/1	10.51/1	14.87/1	9.19/1	-0.60/15
mehg4	1.06/1	5.39/1	5.17/1	-5.73/1	5.8/1	7.07/1	3.89/1	-2.19/1
pcb1	-1.93/0	-1.46/0	0.02/0	0.27/0	2.41/1	-0.26/0	0.35/0	-3.38/1
pcb2	-0.28/0	-0.33/0	0.30/0	0.42/0	2.11/1	1.22/1	0.4/0	-3.22/1
pcb3	-0.07/0	-0.99/0	0.58/0	-0.03/0	1.83/1	2.66/1	0.46/0	-3.06/1

For *sepw1* and *itgb1b.2*, ROBI for negative thresholds was greater than the result for positive thresholds. Therefore, the table shows the transformation with a threshold of -0.5 for *sepw1* and *itgb1b.2*

Table 2 ANOVA results for the thresholds 0.5–5 for the genes from Table 1

	cish	fn1b	socs3a	sepw1	mmp13a	mmp9	zgc:103438	itgb1b.2
0.5	0.0352 **	0.0125 **	0.1443	0.0089 **	0.0089 **	0.3806	0.0125 **	0.1161 *
1	0.0352 **	0.0020 **	0.0125 **	0.0089 **	0.0089 **	0.1681	0.002 **	0.1161 *
1.5	0.0088 **	0.0020 **	0.0020 **	0.0089 **	0.0089 **	0.0812 *	0.002 **	0.1161 *
2	0.0088 **	0.0020 **	0.0020 **	0.0236 **	0.0236 **	0.0154 **	0.002 **	0.1161 *
2.5	0.1161	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0154 **	0.002 **	0.1161 *
3	0.5082	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.002 **	0.1161 *
3.5	1	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.002 **	0.1161 *
4	1	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0317 **	0.1161 *
4.5	1	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0317 **	0.1161 *
5	1	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0020 **	0.0317 **	0.1161 *
ROBI	8	20	18	20	20	15	20	10
Normalized ROBI	0.4	1	0.9	1	1	0.75	1	0.5
Most significant treatment	chlor	mehg	mehg	mehg	mehg	mehg	mehg	pcb

Significant and highly significant *p* values are marked with one or two asterisks and the treatment (chlor, mehg, or pcb) which had the most significant result in the post-hoc test

*significant; **highly significant

and to avoid some statistical pitfalls. An aggregated analysis allows to compare gene expression of all genes for all treatments simultaneously and to identify genes that discriminate between these treatments (DeConde et al. 2006). By the extension of the glm with a Bayesian approach, the intrinsic problems of small sample size and perfect segregation can be taken care of (Gelman et al. 2008).

Nevertheless, problems of multiple testing and the resulting large number of statistical significantly regulated genes still remain. Additionally, data noise can mask the effects of substances on the gene expression (Hallare et al. 2011). In order to compensate this, our approach uses both fold-change cut-offs and binary variables effectively reducing variability within the

data. A problem arising here is to determine which fold-change cut-off value to choose. Several studies showed the dependence of the results of transcriptome analysis on the choice of arbitrary thresholds (e.g., Dalman et al. 2012). Because of the dependency of the gene expression on a multitude of factors, it is unlikely that there is one fold-change applicable to all genes of one gene expression study (Lettieri 2005). We found that discriminative genes can be selected with different thresholds, and these thresholds vary from substance to substance and even gene to gene. When working with fold-changes or algorithms that use fold-changes, one has to keep this arbitrariness of predefined thresholds in mind. Because there is no way to determine which is the best

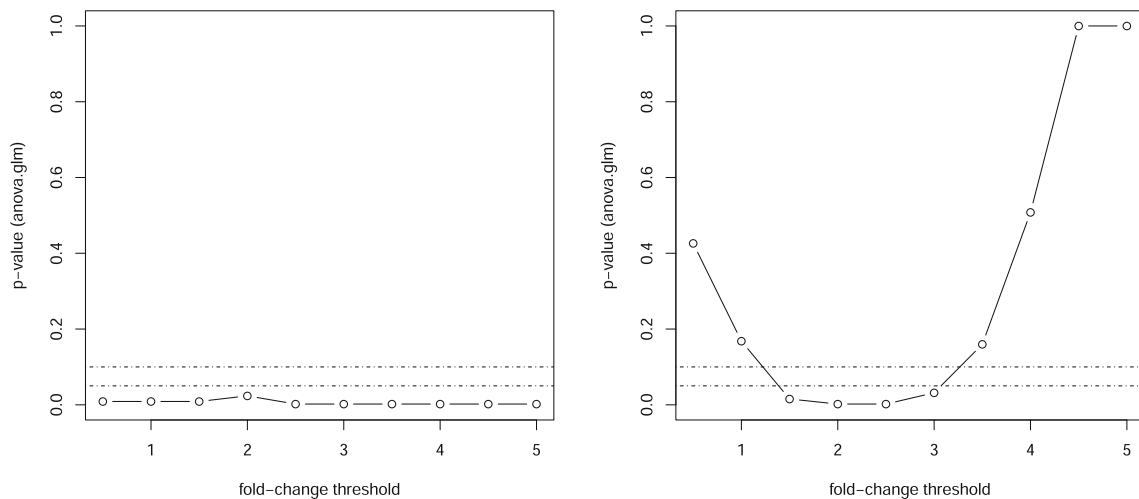
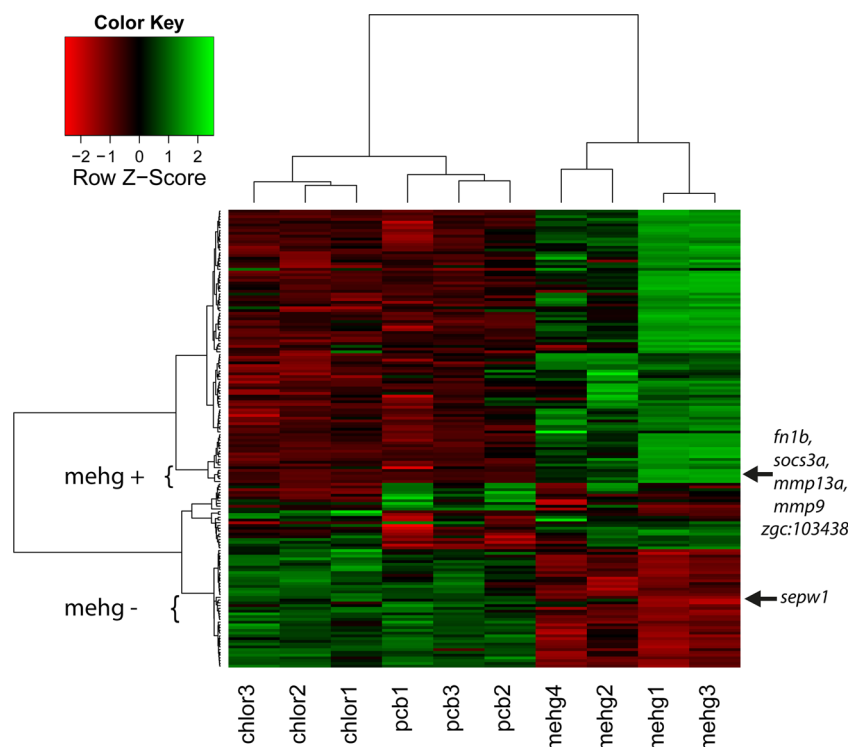


Fig. 2 Significance profiles of two selected genes. *Left:* matrix metalloproteinase 13a (*mmp13a*). *Right:* Keratin 18 (*krt18*). On the *x-axis* are the different applied fold-change thresholds, on the *y-axis*

are the *p* values of anova.glm. *Dashed lines* mark 10 and 5 % significance levels

Fig. 3 Heatmap clustering (Euclidean distance, Ward's clustering) of all genes with a mean fold-change >2 for at least one treatment using the heatmap.2 function from the R-Package gplots (R-Version 3.0.2). Upregulation is colored green, downregulation red (color code in the top left). On the right side, the positions of *fn1b*, *socs3a*, *mmp13a*, *mmp9*, *zgc:103438*, and *sepw1*, which reacted significantly to methylmercury from Table 2 are marked. On the left side, the methylmercury regulation clusters are indicated



threshold before analyzing the transcriptome data, using (uninformed) fold-changes as the only criterion will often fail to distinguish between detecting the effects of no interests and referable to contaminants (Draghici 2002). For fold-change threshold screening, a multi-threshold approach was used with ten different fold-change thresholds (Table 2).

The significance profile in Fig. 2 (left) as well as the ROBI (Table 2) revealed that there is strong evidence that *mmp13a* is a potential indicator gene for methylmercury. Yang et al. (2010) showed that methylmercury suppresses the formation of the tail primordium of *D. rerio*. From the genes they found to be significantly expressed under methylmercury treatment, *mmp9* and *mmp13a* stood out as being strongly upregulated. They concluded that *mmp9* and *mmp13a* are involved in the tail development and were regulated when treated with methylmercury. Ho et al. (2013) examined the changes in gene expression in the central nervous system that were induced by methylmercury. Again, the genes with strong upregulation were *mmp9* and *mmp13a*. Both experiments were confirmed by real-time PCR. By using fold-change threshold screening in our study, both *mmp9* and *mmp13a* were among the most robust and highly significant genes for treatment with methylmercury. In addition, most of the genes that were found to be significant by Ho et al. (2013) also had a high ROBI in our study (e.g., *fn1b*, see Table 2). These results demonstrate that fold-change threshold screening is capable of finding robust and significant genes and providing reproducible results.

The positions of the genes which were selected as potential indicators for methylmercury were marked in the heatmap

(Fig. 3). With the exception of *sepw1*, they all form one very tight cluster which contains genes that are upregulated in the presence of methylmercury (“mehg +” in Fig. 3). This cluster contains almost exclusively genes that were selected by fold-change threshold screening. Thus, the algorithm is capable of reducing the dataset to a very small number of significant genes or gene patterns. The reason why *sepw1* does not belong to this cluster is that it is significantly downregulated in the presence of methylmercury (“mehg -” in Fig. 3). It is part of the cluster at the bottom of the heatmap which contains the genes that show a negative correlation with methylmercury.

Taking the significance profile of *krt18* as an example, it is possible to use the information of every threshold to determine a useful fold-change cut-off for any further analysis like GO-Term analysis (Fig. 2, right). Only when a threshold between 1.5 and 3 was applied, the expression was significantly affected. The other tested thresholds were far less significant.

Upregulation of keratin 18 (*krt18*) in the presence of methylmercury is found in several microarray studies (Nøstbakken et al. 2012; Cuello et al. 2012). In *D. rerio* *krt18*, a member of type I keratin genes, is strongly linked to the development and regeneration of scales as well as the dorsal fin (Padhi et al. 2004; Wang et al. 2006). In the case of our study, neither a too small nor a too big cut-off would have led to a result suggesting that *krt18* is a potential indicator for methylmercury. Only with the help of fold-change threshold screening and the significance profile, it is possible to unmask the effect of methylmercury. Additionally, with the help of the significance profile, it becomes possible to choose fold-change cut-off values

suitable for a certain dataset (or single genes) despite the strong overall data noise.

An additional interesting gene is *cyp1a*, which is known to be an indicator for PCB (Jönsson et al. 2007; Otte et al. 2008). As in our analysis, *cyp1a* was significantly differently regulated in the presence of PCB as well as chlorpyrifos, making it difficult to distinguish between the two of them. However, using cut-off threshold values up to 10 (instead of up to 5) for our algorithm, results of the post-hoc test indicated a stronger effect in the presence of PCB, which is in line with results given in literature.

Conclusions

Because of the large number of tested genes, one problem of transcriptome analysis is multiple testing and adjusting the significance level, for example with false-discovery-rate (FDR). These adjustments always raise the chance for type II errors (Underwood and Chapman 2003). In ecotoxicology, overlooking an impact although it is there might result in collapse of populations and catastrophic outcomes for the ecosystem, which is why in environmental study type II errors are considered to be more costly than type I errors (Mapstone 1995). For screening the adverse effects of a contaminated environmental sample on a model organism, it is therefore in particular important to keep the chance for a type II error low. Therefore, fold-change threshold screening uses a different approach to find discriminative genes without reducing the significance level.

Fold-change threshold screening is an easy to understand yet powerful algorithm especially suited for aggregated analysis of gene expression data. It is able to face an important problem of transcriptome analysis in ecotoxicology, namely to test the robustness of indicated genes while circumventing the need for choosing a more or less arbitrary fold-change cut-off threshold, e.g., functional analysis. By providing a robustness indicator (ROBI), an indication of the uncertainty and indicative power of each gene can be given. This robustness indicator allows reducing data dimensionality to a manageable set of variables that can be used in environmental assessment. These patterns of gene expression (“fingerprints,” “barcodes”) could be used as biomarkers identifying relevant toxicants (Aardema and MacGregor 2002; Yang et al. 2007; Piña and Barata 2011). With robust indicator genes, the full potential of transcriptomics can be used. However, data noise, which is caused by high variability, can often mask the effects of the treatments. To use gene expression data as measurement and predictor for pollutant exposure as well as its impact on ecosystems, it is necessary to unmask these effects. Fold-change threshold screening is capable of unmasking hidden gene expression patterns and revealing appropriate fold-change thresholds for all genes.

As already pointed out, there is no real gold standard for the analysis of gene expression data up to date, but there are a lot of options to choose from. Witten and Tibshirani (2007) have shown that the chosen method will have great impact on the set of genes which is selected. They furthermore conclude that the choice of method should be based on the biological system of interest. Fold-change threshold screening was created especially for aggregated analysis of gene expression data from *D. rerio* in ecotoxicology. Most studies select genes from a huge number of highly significant genes which they know beforehand could be biologically relevant based on functional a priori knowledge. In contrast, our approach is capable of detecting discriminative genes without a priori information needed. Thus, it is a strong tool for dimensionality reduction as well as feature selection, two main goals of statistical pattern recognition.

In our study, we combined microarray data for three sediment-typical contaminants in an aggregated data set, analyzed simultaneously. We propose this approach as a powerful tool to detect contamination of simultaneously occurring pollutants and it can thus be a potential tool in ecotoxicological sediment screening, although it is possible to use it with other transcriptome data.

Acknowledgments The present study was part of the research funding project DanTox (DanTox—a novel joint research project using zebrafish (*Danio rerio*) to identify specific toxicity and molecular modes of action of sediment-bound pollutants). The authors acknowledge the financial support by the German Federal Ministry of Education and Research (BMBF grant 02WU1053). Also, the authors acknowledge the data provision from the GENDarT2 project (BMBF grant AZ:0315190 B).

Conflict of interest The authors declare that they have no competing interests.

References

- Aardema MJ, MacGregor JT (2002) Toxicology and genetic toxicology in the new era of “toxicogenomics”: impact of “-omics” technologies. *Mutat Res Fundam Mol Mech Mutagen* 499:13–25. doi:10.1016/S0027-5107(01)00292-5
- Benjamini Y, Drai D, Elmer G et al (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 125:279–284. doi:10.1016/S0166-4328(01)00297-2
- Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 573:83–92. doi:10.1016/j.febslet.2004.07.055
- Busquet F, Strecker R, Rawlings JM et al (2014) OECD validation study to assess intra- and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. *Regul Toxicol Pharmacol* 69:496–511. doi:10.1016/j.yrtph.2014.05.018
- Chapman PM (2000) The sediment quality triad: then, now and tomorrow. *Int J Environ Pollut* 13:351–356
- Chapman PM, Anderson J (2005) A decision-making framework for sediment contamination. *Integr Environ Assess Manag* 1:163–173. doi:10.1897/2005-013R.1

- Cordero F, Botta M, Calogero RA (2007) Microarray data analysis and mining approaches. *Brief Funct Genomic Proteomic* 6:265–281. doi:10.1093/bfgp/elm034
- Cuello S, Ximénez-Embún P, Ruppen I et al (2012) Analysis of protein expression in developmental toxicity induced by MeHg in zebrafish. *Analyst* 137:5302–5311. doi:10.1039/c2an35913h
- Dalman MR, Deeter A, Nimishakavi G, Duan Z-H (2012) Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinf* 13:S11. doi:10.1186/1471-2105-13-S2-S11
- DeConde RP, Hawley S, Falcon S et al (2006) Combining results of microarray experiments: a rank aggregation approach. *Stat Appl Genet Mol Biol* 5:15. doi:10.2202/1544-6115.1204
- Denslow ND, Garcia-Reyero N, Barber DS (2007) Fish “n” chips: the use of microarrays for aquatic toxicology. *Mol Biosyst* 3:172–177. doi:10.1039/b612802p
- Ding C, Hanchuan P (2005) Minimum redundancy feature selection from microarray gene expression data. *J Bioinforma Comput Biol* 3:185–205
- Draghici S (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov Today* 7:55–63
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* 95:14863–14868. doi:10.1073/pnas.95.25.14863
- Fedorenkova A, Vonk J (2010) Ecotoxicogenomics: bridging the gap between genes and populations. *Environ Sci Technol* 44:4328–4333
- Feiler U, Höss S, Ahlf W et al (2013) Sediment contact tests as a tool for the assessment of sediment quality in German waters. *Environ Toxicol Chem* 32:144–155. doi:10.1002/etc.2024
- Gelman A, Jakulin A, Pittau MG, Su Y-S (2008) A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2:1360–1383. doi:10.1214/08-AOAS191
- Hallare AV, Seiler TB, Hollert H (2011) The versatile, changing, and advancing roles of fish in sediment toxicity assessment—a review. *J Soils Sediments* 11:141–173. doi:10.1007/s11368-010-0302-7
- Hausen J, Otte JC, Yang L et al (2015) Fishing for contaminants: aggregated analysis of gene expression data reveals discriminative genes for effects on *Danio rerio* embryogenesis in response to sediment-typical pollution. *Environ Sci Eur* (in this issue)
- Ho NY, Yang L, Legradi J et al (2013) Gene responses in the central nervous system of zebrafish embryos exposed to the neurotoxicant methyl mercury. *Environ Sci Technol* 47:3316–3325. doi:10.1021/es3050967
- Hollert H, Keiter S, König N et al (2003) A new sediment contact assay to assess particle-bound pollutants using zebrafish (*Danio rerio*) embryos. *J Soils Sediments* 3:197–207. doi:10.1065/jss2003.09.085
- Höss S, Ahlf W, Fahnenstich C et al (2010) Variability of sediment-contact tests in freshwater sediments with low-level anthropogenic contamination—determination of toxicity thresholds. *Environ Pollut* 158:2999–3010. doi:10.1016/j.envpol.2010.05.013
- Jönsson ME, Jenny MJ, Woodin BR et al (2007) Role of AHR2 in the expression of novel cytochrome P450 1 family genes, cell cycle genes, and morphological defects in developing zebra fish exposed to 3,3',4,4',5-pentachlorobiphenyl or 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Sci* 100:180–193. doi:10.1093/toxsci/kfm207
- Keiter S, Peddinghaus S, Feiler U et al (2010) DanTox—a novel joint research project using zebrafish (*Danio rerio*) to identify specific toxicity and molecular modes of action of sediment-bound pollutants. *J Soils Sediments* 10:714–717. doi:10.1007/s11368-010-0221-7
- Keiter SH, Braunbeck T, Feiler U et al (2013) DanTox - Entwicklung und Anwendung eines Verfahrens zur Ermittlung spezifischer Toxizität und molekularer Wirkungsmechanismen sedimentgebundener Umweltschadstoffe mit dem Zebrafisch (*Danio rerio*) : Schlussbericht. Aachen
- Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. *J Comput Biol* 7:819–837. doi:10.1089/10665270050514954
- Kosmehl T, Otte JC, Yang L et al (2012) A combined DNA-microarray and mechanism-specific toxicity approach with zebrafish embryos to investigate the pollution of river sediments. *Reprod Toxicol* 33:245–253. doi:10.1016/j.reprotox.2012.01.005
- Legradi J (2011) Microarray based transcriptomics and the search for biomarker genes in zebrafish. Ruprecht-Karls Universität, Heidelberg
- Lesaffre E, Albert A (1989) Partial separation logistic regression. *J R Stat Soc Ser B* 51:109–116
- Lettieri T (2005) Recent applications of DNA microarray technology to toxicology and ecotoxicology. *Environ Health Perspect* 4–9. doi:10.1289/ehp.8194
- Mapstone BD (1995) Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecol Appl* 5:401–410
- McCarthy DJ, Smyth GK (2009) Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25:765–771. doi:10.1093/bioinformatics/btp053
- Nøstbakken OJ, Goksøyr A, Martin SAM et al (2012) Marine n-3 fatty acids alter the proteomic response to methylmercury in Atlantic salmon kidney (ASK) cells. *Aquat Toxicol* 106–107:65–75. doi:10.1016/j.aquatox.2011.10.008
- Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nat Genet* 32:261–266. doi:10.1038/ng983
- Otte JC, Andersson C, Abrahamson A et al (2008) A bioassay approach to determine the dioxin-like activity in sediment extracts from the Danube River: ethoxyresorufin-O-deethylase induction in gill filaments and liver of three-spined sticklebacks (*Gasterosteus aculeatus* L.). *Environ Int* 34:1176–1184. doi:10.1016/j.envint.2008.05.004
- Padhi BK, Joly L, Tellis P et al (2004) Screen for genes differentially expressed during regeneration of the zebrafish caudal fin. *Dev Dyn* 231:527–541. doi:10.1002/dvdy.20153
- Pavlidis P (2003) Using ANOVA for gene selection from microarray studies of the nervous system. *Methods* 31:282–289. doi:10.1016/S1046-2023(03)00157-9
- Peng H-Y, Jiang C-F, Fang X, Liu J-S (2014) Variable selection for Fisher linear discriminant analysis using the modified sequential backward selection algorithm for the microarray data. *Appl Math Comput* 238:132–140. doi:10.1016/j.amc.2014.03.141
- Piña B, Barata C (2011) A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. *Aquat Toxicol* 105:40–49. doi:10.1016/j.aquatox.2011.06.006
- R Core Team (2014) R: a language and environment for statistical computing. *R Found Stat Comput* 1:409. doi:10.1007/978-3-540-74686-7
- Reboiro-Jato M, Díaz F, Glez-Peña D, Fdez-Riverola F (2014) A novel ensemble of classifiers that use biological relevant gene sets for microarray classification. *Appl Soft Comput* 17:117–126. doi:10.1016/j.asoc.2014.01.002
- Schiwy S, Bräunig J, Alert H et al (2014) A novel contact assay for testing aryl hydrocarbon receptor (AhR)-mediated toxicity of chemicals and whole sediments in zebrafish (*Danio rerio*) embryos. *Environ Sci Pollut Res Int*. doi:10.1007/s11356-014-3185-0
- Snell TW, Brogdon SE, Morgan MB (2003) Gene expression profiling in ecotoxicology. *Ecotoxicology* 12:475–483
- Strähle U, Scholz S, Geisler R et al (2012) Zebrafish embryos as an alternative to animal experiments—a commentary on the definition of the onset of protected life stages in animal welfare regulations. *Reprod Toxicol* 33:128–132. doi:10.1016/j.reprotox.2011.06.121
- Underwood AJ, Chapman MG (2003) Power, precaution, type II error and sampling design in assessment of environmental

- impacts. *J Exp Mar Biol Ecol* 296:49–70. doi:[10.1016/S0022-0981\(03\)00304-6](https://doi.org/10.1016/S0022-0981(03)00304-6)
- Wang Y-H, Chen Y-H, Wu T-N et al (2006) A keratin 18 transgenic zebrafish Tg(k18(2.9):RFP) treated with inorganic arsenite reveals visible overproliferation of epithelial cells. *Toxicol Lett* 163:191–197. doi:[10.1016/j.toxlet.2005.10.024](https://doi.org/10.1016/j.toxlet.2005.10.024)
- Witten DM, Tibshirani R (2007) A comparison of fold-change and the t-statistic for microarray data analysis. Stanford Univ. Tech. Rep
- Yang IV, Chen E, Hasseman JP et al (2002) Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol* 3:1–12
- Yang L, Kemadjou JR, Zinsmeister C et al (2007) Transcriptional profiling reveals barcode-like toxicogenomic responses in the zebrafish embryo. *Genome Biol* 8:R227. doi:[10.1186/gb-2007-8-10-r227](https://doi.org/10.1186/gb-2007-8-10-r227)
- Yang L, Ho NY, Müller F, Strähle U (2010) Methyl mercury suppresses the formation of the tail primordium in developing zebrafish embryos. *Toxicol Sci* 115:379–390. doi:[10.1093/toxsci/kfq053](https://doi.org/10.1093/toxsci/kfq053)
- Zhang S, Cao J (2009) A close examination of double filtering with fold change and T test in microarray analysis. *BMC Bioinf* 10:402. doi:[10.1186/1471-2105-10-402](https://doi.org/10.1186/1471-2105-10-402)
- Zorn C (2005) A solution to separation in binary response models. *Polit Anal* 13:157–170. doi:[10.1093/pan/mpi009](https://doi.org/10.1093/pan/mpi009)